

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07H 21/02, 21/04, C07K 1/00, 14/435, C12N 15/12, 15/63, 15/64, C12P 21/00	A1	(11) International Publication Number: WO 97/44347 (43) International Publication Date: 27 November 1997 (27.11.97)
(21) International Application Number: PCT/US96/07477 (22) International Filing Date: 21 May 1996 (21.05.96) (71) Applicants (for all designated States except US): SMITHK-LINE BEECHAM CORPORATION [US/US]; Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US). PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): DUNNINGTON, Damien, D. [GB/US]; 23 Forsythia Court, Lafayette Hill, PA 19444 (US). FRANTZ, James, D. [IS/US]; 22 Jenison Street, Newton, MA 02169 (US). SHOELSON, Steven, E. [US/US]; 8 Mocassin Path, Natick, MA 01760 (US). (74) Agents: BAUMEISTER, Kirk et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).		(81) Designated States: AL, AM, AU, BB, BG, BR, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, LK, LR, LT, LV, MD, MG, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>
(54) Title: HUMAN CIS PROTEIN (57) Abstract Isolated nucleic acid encoding a human cytokine-inducible SH2-containing protein, protein obtainable from the nucleic acid, recombinant host cells transformed with the nucleic acid and use of the protein and nucleic acid sequence are disclosed.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HUMAN CIS PROTEIN

Field of the Invention

The present invention relates to an isolated human
5 cytokine-inducible SH2-containing (CIS) gene; to
essentially pure human CIS protein; and to compositions
and methods of producing and using human CIS sequences
and proteins.

Background of the Invention

A number of polypeptide growth factors and hormones
mediate their cellular effects through a signal
transduction pathway. Transduction of signals from the
cell surface receptors for these ligands to
15 intracellular effectors frequently involves
phosphorylation or dephosphorylation of specific protein
substrates by regulatory protein tyrosine kinases (PTK)
and phosphatases. Tyrosine phosphorylation is a major
mediator of signal transduction in multicellular
20 organisms. Receptor-bound, membrane-bound and
intracellular PTKs regulate cell proliferation, cell
differentiation and signalling processes in
hematopoietic cells.

Aberrant protein tyrosine kinase activity has been
25 implicated or is suspected in a number of pathologies
such as diabetes, atherosclerosis, psoriasis, septic
shock, bone loss, anemia, many cancers and other
proliferative diseases. Accordingly, tyrosine kinases
and the signal transduction pathways which they are part
30 of are potential targets for drug design. For a review,
see Levitzki et al. in *Science* 267, 1782-1788 (1995).

Many of the proteins comprising signal transduction
pathways are present at low levels and often have
opposing activities. The properties of these signalling
35 molecules allow the cell to control transduction by
means of the subcellular location and juxtaposition of
effectors as well as by balancing activation with

repression such that a small change in one pathway can achieve a switching effect.

The formation of transducing complexes by juxtaposition of the signalling molecules through protein-protein interactions are mediated by specific docking domain sequence motifs. Src homology 2 (SH2) domains, which are conserved non-catalytic sequences of approximately 100 amino acids found in a variety of signalling molecules such as non-receptor PTKs and kinase target effector molecules and in oncogenic proteins, play a critical role. The SH2 domains are highly specific for short phosphotyrosine-containing peptide sequences found in autophosphorylated PTK receptors or intracellular tyrosine kinases.

One approach towards the pharmacological regulation of signal transduction pathways is to design inhibitory ligands which selectively bind to a chosen SH2 domain and thus block the interaction of a phosphorylated protein tyrosine kinase with its SH2-containing target molecule, thereby disrupting signal transduction. Any selective inhibitors would provide a useful lead for drug development.

Cytokine-inducible SH2-containing protein, otherwise known as CIS or SIC, is an SH2 domain containing protein identified in the mouse as an early response gene induced by certain cytokines such as interleukins 2 and 3, granulocyte-macrophage colony stimulating factor and erythropoietin (EPO). See Yoshimura et al., *EMBO J.* 14, 2816-2826 (1995). CIS is expressed in liver, kidney, heart, stomach and lung tissues. It binds to the tyrosine-phosphorylated IL3 or EPO receptors and when overexpressed, inhibits signal transduction through these receptors. CIS appears to belong to the "adaptor" class of SH2-containing proteins and may function by recruitment of negative regulators of signaling such as phosphatases or by masking binding sites for positive effectors. Inactivation of CIS may

be expected to enhance signaling through the IL-3 or EPO receptors, thereby up-regulating the effects of these cytokines. In the case of EPO, such up-regulation may have utility as a means of stimulating hematopoiesis.

5 Therefore, specific inhibitors of CIS may be useful in the treatment of anemia.

Binding of CIS to cytokine receptors is mediated by SH2-phosphotyrosine interactions, which are amenable to disruption by small molecule agents. Discovery of such
10 agents is best carried out using the human CIS molecule or fragment thereof, and an appropriate ligand such as the tyrosine-phosphorylated EPO receptor or a synthetic phosphopeptide. Thus, a need exists for provision of the nucleotide and amino acid sequences corresponding to
15 human CIS, for compounds which modulate the activity of CIS homologs and isoforms, for methods to identify such modulators and for reagents useful in such methods.

Summary of the Invention

20 Accordingly, one aspect of the present invention is an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide encoding human CIS having the nucleotide sequence as set forth in SEQ ID NO:1 from
25 nucleotide 72 to 846;

(b) a polynucleotide capable of hybridizing to the complement of a polynucleotide according to (a) under moderately stringent hybridization conditions and which encodes a functional human CIS; and

30 (c) a degenerate polynucleotide according to (a) or (b).

Another aspect of the invention is a functional polypeptide encoded by the polynucleotides of the invention.

35 Another aspect of the invention is a method for preparing essentially pure human CIS protein comprising culturing a recombinant host cell comprising a vector

comprising a polynucleotide of the invention under conditions promoting expression of the protein and recovery thereof.

Another aspect of the invention is an antisense
5 oligonucleotide comprising a sequence which is capable of binding to the polynucleotide of the invention.

Another aspect of the invention is a modulator of the polypeptides of the invention.

Another aspect of the invention is a method for
10 assaying a medium for the presence of a substance that modulates CIS activity comprising the steps of:

(a) providing a CIS protein having the amino acid sequence of CIS (SEQ ID NO:2) or a functional derivative thereof and a cellular binding partner or synthetic
15 analog thereof;

(b) incubating with a test substance which is suspected of modulating CIS activity under conditions which permit the formation of a CIS protein/cellular binding partner complex;

20 (c) assaying for the presence of the complex, free CIS protein or free cellular binding partner; and

(d) comparing to a control to determine the effect of the substance.

Another aspect of the invention is a method for
25 assaying for the presence of a substance that modulates CIS activity by direct binding to CIS protein comprising the steps of:

(a) providing a labelled CIS protein having the amino acid sequence of CIS (SEQ ID NO:2) or a
30 functional derivative thereof;

(b) providing solid support-associated modulator candidates;

(c) incubating a mixture of the labelled CIS protein with the support-associated modulator candidates
35 under conditions which can permit the formation of a CIS protein/modulator candidate complex;

(d) separating the solid support from free soluble labelled CIS protein;

(e) assaying for the presence of solid support-associated labelled protein;

5 (f) isolating the solid support complexed with labelled CIS protein; and

(g) identifying the modulator candidate.

Another aspect of the invention is CIS protein modulating compounds identified by the methods of the
10 invention.

Another aspect of the invention is a method for the treatment of a patient having need to modulate CIS activity comprising administering to the patient a therapeutically effective amount of the modulating
15 compounds of the invention.

Brief Description of the Drawing

Figure 1 is an amino acid sequence alignment of human CIS with murine CIS.

20

Detailed Description of the Invention

As used herein, the term "CIS gene" refers to DNA molecules comprising a nucleotide sequence that encodes human CIS. The CIS gene sequence is listed in SEQ ID
25 NO:1. The coding region of the CIS gene consists of nucleotides 72-846 of SEQ ID NO:1. The deduced 258 amino acid sequence of the gene product CIS is listed in SEQ ID NO:2.

As used herein, the term "functional fragments"
30 when used to modify a specific gene or gene product means a less than full length portion of the gene or gene product which retains substantially all of the biological function associated with the full length gene or gene product to which it relates. An example of a
35 functional fragment of human CIS is the isolated SH2 domain lacking flanking sequences. To determine whether a fragment of a particular gene or gene product is a

functional fragment, fragments are generated by well-known nucleolytic or proteolytic techniques or by the polymerase chain reaction and the fragments tested for the described biological function.

5 As used herein, an "antigen" refers to a molecule containing one or more epitopes that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is also used herein interchangeably with "immunogen."

10 As used herein, the term "epitope" refers to the site on an antigen or hapten to which a specific antibody molecule binds. The term is also used herein interchangeably with "antigenic determinant" or "antigenic determinant site."

15 As used herein, "monoclonal antibody" is understood to include antibodies derived from one species (e.g., murine, rabbit, goat, rat, human, etc.) as well as antibodies derived from two (or perhaps more) species (e.g., chimeric and humanized antibodies).

20 As used herein, a coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences.

25 The coding sequences need not be contiguous to one another so long as the expressed sequence is ultimately processed to produce the desired protein.

 As used herein, "recombinant" polypeptides refer to polypeptides produced by recombinant DNA techniques;
30 i.e., produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide. "Synthetic" polypeptides are those prepared by chemical synthesis.

As used herein, a "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

5 As used herein, a "vector" is a replicon, such as a plasmid, phage, or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

10 As used herein, a "reference" gene refers to the wild type human CIS gene sequence of the invention and is understood to include the various sequence polymorphisms that exist, wherein nucleotide substitutions in the gene sequence exist, but do not affect the essential function of the gene product.

15 As used herein, a "mutant" gene refers human CIS sequences different from the reference gene wherein nucleotide substitutions and/or deletions and/or insertions result in perturbation of the essential function of the gene product.

20 As used herein, a DNA "coding sequence of" or a "nucleotide sequence encoding" a particular protein, is a DNA sequence which is transcribed and translated into a polypeptide when placed under the control of appropriate regulatory sequences.

25 As used herein, a "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bound at
30 its 3' terminus by a translation start codon (e.g., ATG) of a coding sequence and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be
35 found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for

the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

As used herein, DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers and the like, which collectively provide for the expression (i.e., the transcription and translation) of a coding sequence in a host cell.

As used herein, a control sequence "directs the expression" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

As used herein, a "host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous DNA sequence.

As used herein, a cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes and yeasts, for example, the exogenous DNA may be maintained on an episomal element, such as a plasmid. With respect to eukaryotic cells, a stably transformed or transfected cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the exogenous DNA.

As used herein, "transfection" or "transfected" refers to a process by which cells take up foreign DNA and integrate that foreign DNA into their chromosome. Transfection can be accomplished, for example, by various techniques in which cells take up DNA (e.g., calcium phosphate precipitation, electroporation, assimilation of liposomes, etc.) or by infection, in which viruses are used to transfer DNA into cells.

As used herein, a "target cell" is a cell that is selectively transfected over other cell types (or cell lines).

As used herein, a "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

As used herein, a "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature. Thus, when the heterologous region encodes a gene, the gene will usually be flanked by DNA that does not flank the gene in the genome of the source animal. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Allelic variation or naturally occurring mutational events do not give rise to a heterologous region of DNA, as used herein.

As used herein, a "modulator" of a polypeptide is a substance which can affect the polypeptide function.

An aspect of the present invention is isolated polynucleotides encoding a human CIS protein and substantially similar sequences. Isolated polynucleotide sequences are substantially similar if they are capable of hybridizing under moderately stringent conditions to SEQ ID NO:1 or they encode DNA

sequences which are degenerate to SEQ ID NO:1 or are degenerate to those sequences capable of hybridizing under moderately stringent conditions to SEQ ID NO:1.

Moderately stringent conditions is a term understood by the skilled artisan and has been described in, for example, Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd edition, Vol. 1, pp. 101-104, Cold Spring Harbor Laboratory Press (1989). An exemplary hybridization protocol using moderately stringent conditions is as follows. Nitrocellulose filters are prehybridized at 65°C in a solution containing 6X SSPE, 5X Denhardt's solution (10g Ficoll, 10g BSA and 10g polyvinylpyrrolidone per liter solution), 0.05% SDS and 100 ug/ml tRNA. Hybridization probes are labeled, preferably radiolabelled (e.g., using the Bios TAG-IT[®] kit). Hybridization is then carried out for approximately 18 hours at 65°C. The filters are then washed twice in a solution of 2X SSC and 0.5% SDS at room temperature for 15 minutes. Subsequently, the filters are washed at 58°C, air-dried and exposed to X-ray film overnight at -70°C with an intensifying screen.

Degenerate DNA sequences encode the same amino acid sequence as SEQ ID NO:2 or the proteins encoded by that sequence capable of hybridizing under moderately stringent conditions to SEQ ID NO:1, but have variation(s) in the nucleotide coding sequences because of the degeneracy of the genetic code. For example, the degenerate codons UUC and UUU both code for the amino acid phenylalanine, whereas the four codons GGX all code for glycine.

Alternatively, substantially similar sequences are defined as those sequences in which about 66%, preferably about 75% and most preferably about 90%, of the nucleotides or amino acids match over a defined length of the molecule. As used herein, substantially similar refers to the sequences having similar identity

to the sequences of the instant invention. Thus nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Protein sequences that are substantially the same can be identified by techniques such as proteolytic digestion, gel electrophoresis and/or microsequencing. Excluded from the definition of substantially similar sequences is the murine CIS gene.

Embodiments of the isolated polynucleotides of the invention include DNA, genomic DNA and RNA, preferably of human origin. A method for isolating a nucleic acid molecule encoding a CIS protein is to probe a genomic or cDNA library with a natural or artificially designed probe using art recognized procedures. See, e.g., "Current Protocols in Molecular Biology", Ausubel et al. (eds.) Greene Publishing Association and John Wiley Interscience, New York, 1989, 1992. The ordinarily skilled artisan will appreciate that SEQ ID NO:1 or fragments thereof comprising at least 15 contiguous nucleotides are particularly useful probes. It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes would enable the ordinarily skilled artisan to isolate complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding CIS proteins from human, mammalian or other animal sources or to screen such sources for related sequences, e.g., additional members of the family, type and/or subtype, including transcriptional regulatory and control elements as well as other stability, processing, translation and tissue specificity-determining regions from 5' and/or 3' regions relative to the coding sequences disclosed herein, all without undue experimentation.

Another aspect of the invention is functional polypeptides encoded by the polynucleotides of the invention. An embodiment of a functional polypeptide of the invention is the human CIS protein having the amino acid sequence set forth in SEQ ID NO:2.

Another aspect of the invention is a method for preparing essentially pure human CIS protein. Yet another aspect is the human CIS protein produced by the preparation method of the invention. This protein has the amino acid sequence listed in SEQ ID NO:2 and includes variants with a substantially similar amino acid sequence that have the same function. The proteins of this invention are preferably made by recombinant genetic engineering techniques by culturing a recombinant host cell containing a vector encoding the polynucleotides of the invention under conditions promoting the expression of the protein and recovery thereof.

The isolated polynucleotides, particularly the DNAs, can be introduced into expression vectors by operatively linking the DNA to the necessary expression control regions, e.g., regulatory regions, required for gene expression. The vectors can be introduced into an appropriate host cell such as a prokaryotic, e.g., bacterial, or eukaryotic, e.g., yeast or mammalian cell by methods well known in the art. See Ausubel et al., *supra*. The coding sequences for the desired proteins, having been prepared or isolated, can be cloned into any suitable vector or replicon. Numerous cloning vectors are known to those of skill in the art and the selection of an appropriate cloning vector is a matter of choice. Examples of recombinant DNA vectors for cloning and host cells which they can transform include, but are not limited to, the bacteriophage (*E. coli*), pBR322 (*E. coli*), pACYC177 (*E. coli*), pKT230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-*E. coli* gram-negative bacteria), pHV14 (*E. coli* and *Bacillus*

subtilis), pBD9 (*Bacillus*), pIJ61 (*Streptomyces*), pUC6 (*Streptomyces*), YIp5 (*Saccharomyces*), a baculovirus insect cell system, a *Drosophila* insect system, YCp19 (*Saccharomyces*) and pSV2neo (mammalian cells). See
5 generally, "DNA Cloning": Vols. I & II, Glover et al. ed. IRL Press Oxford (1985) (1987); and T. Maniatis et al. ("Molecular Cloning" Cold Spring Harbor Laboratory (1982)).

The gene can be placed under the control of control
10 elements such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired protein is transcribed into RNA in the host cell transformed by a vector containing the expression construct. The coding
15 sequence may or may not contain a signal peptide or leader sequence. The proteins of the present invention can be expressed using, for example, the *E. coli* tac promoter or the protein A gene (*spa*) promoter and signal sequence. Leader sequences can be removed by the
20 bacterial host in post-translational processing. See, e.g., U.S. Patent Nos. 4,431,739; 4,425,437 and 4,338,397.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for
25 regulation of the expression of the protein sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art. Exemplary are those which cause the expression of a gene to be turned on or off in response to a chemical or
30 physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

35 An expression vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning

and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences, i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. Modification of the sequences encoding the particular antigen of interest may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector, such as the cloning vectors described above. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site.

In some cases, it may be desirable to produce mutants or analogues of human CIS protein. Mutants or analogues may be prepared by the deletion of a portion of the sequence encoding the protein, by insertion of a sequence, and/or by substitution of one or more nucleotides within the sequence. Techniques for modifying nucleotide sequences, such as site-directed mutagenesis, are well known to those skilled in the art. See, e.g., T. Maniatis et al., *supra*; "DNA Cloning," Vols. I and II, *supra*; and "Nucleic Acid Hybridization", *supra*.

Depending on the expression system and host selected, the proteins of the present invention are produced by growing host cells transformed by an expression vector described above under conditions whereby the protein of interest is expressed. Preferred mammalian cells include human embryonic kidney cells (293), monkey kidney cells, fibroblast (COS) cells, Chinese hamster ovary (CHO) cells, *Drosophila* or murine

L-cells. If the expression system secretes the protein into growth media, the protein can be purified directly from the media. If the protein is not secreted, it is isolated from cell lysates or recovered from the cell
5 membrane fraction. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

An alternative method to identify proteins of the present invention is by constructing gene libraries,
10 using the resulting clones to transform *E. coli* and pooling and screening individual colonies using polyclonal serum or monoclonal antibodies to human CIS.

The proteins of the present invention may also be produced by chemical synthesis such as solid phase
15 peptide synthesis on an automated peptide synthesizer, using known amino acid sequences or amino acid sequences derived from the DNA sequence of the genes of interest. Such methods are known to those skilled in the art.

The proteins of the present invention or their
20 fragments comprising at least one epitope can be used to produce antibodies, both polyclonal and monoclonal, directed to epitopes corresponding to amino acid sequences disclosed herein. If polyclonal antibodies are desired, a selected mammal such as a mouse, rabbit,
25 goat or horse is immunized with a protein of the present invention, or its fragment, or a mutant protein. Serum from the immunized animal is collected and treated according to known procedures. Serum polyclonal antibodies can be purified by immunoaffinity
30 chromatography or other known procedures.

Monoclonal antibodies to the proteins of the present invention, and to the fragments thereof, can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by
35 using hybridoma technology is well known. Immortal antibody-producing cell lines can be created by cell fusion and also by other techniques such as direct

transformation of B lymphocytes with oncogenic DNA or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al., "Hybridoma Techniques" (1980); Hammerling et al., "Monoclonal Antibodies and T-cell Hybridomas" (1981); Kennett et al., "Monoclonal Antibodies" (1980); and U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,452,570; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against the antigen of interest, or fragment thereof, can be screened for various properties, i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies are useful in purification, using immunoaffinity techniques, of the individual antigens which they are directed against. Alternatively, genes encoding the monoclonals of interest may be isolated from the hybridomas by PCR techniques known in the art and cloned and expressed in the appropriate vectors. The antibodies of this invention, whether polyclonal or monoclonal have additional utility in that they may be employed as reagents in immunoassays, RIA, ELISA, and the like. The antibodies of the invention can be labeled with an analytically detectable reagent such as a radioisotope, fluorescent molecule or enzyme.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, e.g., Liu et al., *Proc. Natl Acad. Sci. USA*, 84, 3439 (1987)), may also be used in assays or therapeutically. Preferably, a therapeutic monoclonal antibody would be "humanized" as described in Jones et al., *Nature*, 321, 522 (1986); Verhoeyen et al., *Science*, 239, 1534 (1988); Kabat et al., *J. Immunol.*, 147, 1709 (1991); Queen et al., *Proc. Natl Acad. Sci. USA*, 86, 10029 (1989); Gorman et al., *Proc. Natl Acad. Sci. USA*, 88, 34181 (1991); and Hodgson et al., *Bio/Technology*, 9:, 421 (1991).

Another aspect of the present invention is modulators of the polypeptides of the invention. Functional modulation of CIS by a substance includes partial to complete inhibition of function, identical
5 function, as well as enhancement of function. Embodiments of modulators of the invention include peptides, oligonucleotides and small organic molecules including peptidomimetics.

Another aspect of the invention is antisense
10 oligonucleotides comprising a sequence which is capable of binding to the polynucleotides of the invention. Synthetic oligonucleotides or related antisense chemical structural analogs can be designed to recognize, specifically bind to and prevent transcription of a
15 target nucleic acid encoding CIS protein by those of ordinary skill in the art. See generally, Cohen, J.S., *Trends in Pharm. Sci.*, 10, 435(1989) and Weintraub, H.M., *Scientific American*, January (1990) at page 40.

Another aspect of the invention is a method for
20 assaying a medium for the presence of a substance that modulates CIS protein function by affecting the binding of CIS protein to cellular binding partners. Examples of modulators include, but are not limited to peptides and small organic molecules including peptidomimetics. A
25 CIS protein is provided having the amino acid sequence of human CIS (SEQ ID NO:2) or a functional derivative thereof together with a cellular binding partner or synthetic analog thereof. The mixture is incubated with a test substance which is suspected of modulating CIS
30 activity, under conditions which permit the formation of a CIS gene product/cellular binding partner complex. An assay is performed for the presence of the complex, free CIS protein or free cellular binding partner and the result compared to a control to determine the effect of
35 the test substance.

Another aspect of the invention is a method for assaying for the presence of a substance that modulates CIS activity by direct binding to CIS protein. Examples of modulators include, but are not limited to, peptides and small organic molecules including peptidomimetics. Modulator candidates are synthesized on a solid support by techniques such as those disclosed in Lam et al., *Nature* 354, 82 (1991) or Burbaum et al., *Proc. Natl. Acad. Sci. USA* 92, 6027 (1995) to provide solid support-associated modulator candidates. A labelled CIS protein is provided having the amino acid sequence of human CIS (SEQ ID NO:2) or a functional derivative thereof. Exemplary labels include directly attached fluorescent or colored dyes, biotin, radioisotopes or epitope tags, which are detectable by a suitable antibody. A mixture of solid support-associated modulator candidates and labelled CIS protein is incubated under conditions which can permit the formation of a CIS protein/modulator candidate complex. The solid support is separated from free soluble labelled CIS protein. An assay is performed for the presence of solid support-associated labelled protein. Solid supports complexed with labelled protein are isolated and the identity of the modulator candidate determined by techniques well known to those skilled in the art, such as the TOF-SIMS method in Brummel et al., *Science* 264, 399-402(1994).

Modulation of CIS function would be expected to have effects on hematopoietic function. Any antagonist modulators so identified would be expected to have up-regulatory effects on the cytokines IL-3 or EPO and be useful as a therapeutic for the treatment and prevention of anemia.

Further, CIS could be used to isolate proteins which interact with it and this interaction could be a target for interference. Inhibitors of protein-protein interactions between CIS and other factors could lead to

the development of pharmaceutical agents for the modulation of CIS activity.

Methods to assay for protein-protein interactions, such as that of a CIS gene product/binding partner complex, and to isolate proteins interacting with CIS are known to those skilled in the art. Use of the methods discussed below enable one of ordinary skill in the art to accomplish these aims without undue experimentation.

10 The yeast two-hybrid system provides methods for detecting the interaction between a first test protein and a second test protein, *in vivo*, using reconstitution of the activity of a transcriptional activator. The method is disclosed in U.S. Patent No. 5,283,173; reagents are available from Clontech and Stratagene. Briefly, CIS cDNA is fused to a Gal4 transcription factor DNA binding domain and expressed in yeast cells. cDNA library members obtained from cells of interest are fused to a transactivation domain of Gal4. cDNA clones which express proteins which can interact with CIS will lead to reconstitution of Gal4 activity and transactivation of expression of a reporter gene such as *Gall-lacZ*. Optionally, the host cells can be co-transfected with a protein tyrosine kinase to induce tyrosine phosphorylation of members of the cDNA library. Such phosphorylation is necessary for optimum interaction with the SH2 domain of CIS.

 An alternative method is screening of λ gt11, λ ZAP (Stratagene) or equivalent cDNA expression libraries with recombinant CIS. Recombinant CIS protein or fragments thereof are fused to small peptide tags such as FLAG, HSV or GST. The peptide tags can possess convenient phosphorylation sites for a kinase such as heart muscle creatine kinase or they can be biotinylated. Recombinant CIS can be phosphorylated with 32 [P] or used unlabeled and detected with streptavidin or antibodies against the tags. λ gt11cDNA

expression libraries are made from cells of interest and are incubated with the recombinant CIS, washed and cDNA clones isolated which interact with CIS. See, e.g., T. Maniatis et al, *supra*.

5 Another method is the screening of a mammalian expression library in which the cDNAs are cloned into a vector between a mammalian promoter and polyadenylation site and transiently transfected in COS or 293 cells followed by detection of the binding protein 48 hours
10 later by incubation of fixed and washed cells with a labelled CIS, preferably iodinated, and detection of bound CIS by autoradiography (See Sims et al., *Science* 241, 585-589 (1988) and McMahon et al., *EMBO J.* 10, 2821-2832 (1991)). In this manner, pools of cDNAs
15 containing the cDNA encoding the binding protein of interest can be selected and the cDNA of interest can be isolated by further subdivision of each pool followed by cycles of transient transfection, binding and autoradiography. Alternatively, the cDNA of interest
20 can be isolated by transfecting the entire cDNA library into mammalian cells and panning the cells on a dish containing CIS bound to the plate. Cells which attach after washing are lysed and the plasmid DNA isolated, amplified in bacteria, and the cycle of transfection and
25 panning repeated until a single cDNA clone is obtained (See Seed et al, *Proc. Natl. Acad. Sci. USA* 84, 3365 (1987) and Aruffo et al., *EMBO J.* 6, 3313 (1987)). If the binding protein is secreted, its cDNA can be obtained by a similar pooling strategy once a binding or
30 neutralizing assay has been established for assaying supernatants from transiently transfected cells. General methods for screening supernatants are disclosed in Wong et al., *Science* 228, 810-815 (1985).

35 Another alternative method is isolation of proteins interacting with CIS directly from cells. Fusion proteins of CIS with GST or small peptide tags are made and immobilized on beads. Biosynthetically labeled or

unlabeled protein extracts from the cells of interest are prepared, incubated with the beads and washed with buffer. Proteins interacting with CIS are eluted specifically from the beads and analyzed by SDS-PAGE.

- 5 Binding partner primary amino acid sequence data are obtained by microsequencing. Optionally, the cells can be treated with agents that induce a functional response such as tyrosine phosphorylation of cellular proteins. An example of such an agent would be a growth factor or
10 cytokine such as erythropoietin or interleukin-3.

- Another alternative method is immunoaffinity purification. Recombinant CIS is incubated with labeled or unlabeled cell extracts and immunoprecipitated with anti-CIS antibodies. The immunoprecipitate is recovered
15 with protein A-Sepharose and analyzed by SDS-PAGE. Unlabelled proteins are labeled by biotinylation and detected on SDS gels with streptavidin. Binding partner proteins are analyzed by microsequencing. Further, standard biochemical purification steps known to those
20 skilled in the art may be used prior to microsequencing.

- Yet another alternative method is screening of peptide libraries for binding partners. Recombinant tagged or labeled CIS is used to select peptides from a peptide or phosphopeptide library which interact with
25 CIS. Sequencing of the peptides leads to identification of consensus peptide sequences which might be found in interacting proteins.

- CIS binding partners identified by any of these methods or other methods which would be known to those
30 of ordinary skill in the art as well as those putative binding partners discussed above can be used in the assay method of the invention. Assaying for the presence of CIS/binding partner complex are accomplished by, for example, the yeast two-hybrid system, ELISA or
35 immunoassays using antibodies specific for the complex. In the presence of test substances which interrupt or inhibit formation of CIS/binding partner interaction, a

decreased amount of complex will be determined relative to a control lacking the test substance.

Assays for free CIS or binding partner are accomplished by, for example, ELISA or immunoassay using
5 specific antibodies or by incubation of radiolabeled CIS with cells or cell membranes followed by centrifugation or filter separation steps. In the presence of test substances which interrupt or inhibit formation of
10 free CIS or free binding partner will be determined relative to a control lacking the test substance.

Another aspect of the invention is pharmaceutical compositions comprising an effective amount of a CIS modulator of the invention and a pharmaceutically
15 acceptable carrier. Pharmaceutical compositions of modulators of this invention for parenteral administration, i.e., subcutaneously, intramuscularly or intravenously or oral administration can be prepared.

The compositions for parenteral administration will
20 commonly comprise a solution of the modulators of the invention or a cocktail thereof dissolved in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be employed, e.g., water, buffered water, 0.4% saline, 0.3% glycine and the
25 like. These solutions are sterile and generally free of particulate matter. These solutions may be sterilized by conventional, well-known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate
30 physiological conditions such as pH adjusting and buffering agents, etc. The concentration of the modulator of the invention in such pharmaceutical formulation can vary widely, i.e., from less than about 0.5%, usually at or at least about 1% to as much as 15
35 or 20% by weight and will be selected primarily based on fluid volumes, viscosities, etc. according to the particular mode of administration selected.

Thus, a pharmaceutical composition of the modulator of the invention for intramuscular injection could be prepared to contain 1 mL sterile buffered water, and 50 mg of a protein of the invention. Similarly, a
5 pharmaceutical composition of the modulator of the invention for intravenous infusion could be made up to contain 250 mL of sterile Ringer's solution, and 150 mg of a modulator of the invention. Actual methods for preparing parenterally administrable compositions are
10 well known or will be apparent to those skilled in the art and are described in more detail in, for example, *Remington's Pharmaceutical Science*, 15th ed., Mack Publishing Company, Easton, Pennsylvania.

The physician will determine the dosage of the
15 present therapeutic agents which will be most suitable and it will vary with the form of administration and the particular compound chosen, and furthermore, it will vary with the particular patient under treatment. Generally, the physician will wish to initiate treatment
20 with small dosages substantially less than the optimum dose of the compound and increase the dosage by small increments until the optimum effect under the circumstances is reached. It will generally be found that when the composition is administered orally, larger
25 quantities of the active agent will be required to produce the same effect as a smaller quantity given parenterally. The therapeutic dosage will generally be from 1 to 10 milligrams per day and higher although it may be administered in several different dosage units.

30 Depending on the patient condition, the pharmaceutical composition of the invention can be administered for prophylactic and/or therapeutic treatments. In therapeutic application, compositions are administered to a patient already suffering from a
35 disease in an amount sufficient to cure or at least partially arrest the disease and its complications. In prophylactic applications, compositions containing the

present compounds or a cocktail thereof are administered to a patient not already in a disease state to enhance the patient's resistance to the disease.

Single or multiple administrations of the
5 pharmaceutical compositions can be carried out with dose levels and pattern being selected by the treating physician. In any event, the pharmaceutical composition of the invention should provide a quantity of the modulators of the invention sufficient to effectively
10 treat the patient.

Additionally, some diseases result from inherited defective genes. These genes can be detected by comparing the sequence of the defective gene with that of a normal one. Individuals carrying mutations in the
15 CIS gene may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis (genomic DNA, mRNA, etc.) may be obtained from a patient's cells, such as from blood, urine, saliva or tissue biopsy, e.g., chorionic villi sampling or removal
20 of amniotic fluid cells and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), etc. prior to analysis. See, e.g., Saiki et al.,
25 *Nature*, 324, 163-166 (1986), Bej, et al., *Crit. Rev. Biochem. Molec. Biol.*, 26, 301-334 (1991), Birkenmeyer et al., *J. Virol. Meth.*, 35, 117-126 (1991), Van Brunt, J., *Bio/Technology*, 8, 291-294 (1990)). RNA or cDNA may also be used for the same purpose. As an example, PCR
30 primers complementary to the nucleic acid of the instant invention can be used to identify and analyze CIS mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal CIS genotype. Point mutations
35 can be identified by hybridizing amplified DNA to radiolabeled CIS RNA of the invention or alternatively, radiolabelled CIS antisense DNA sequences of the

invention. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures (T_m). Such a diagnostic would be particularly useful
5 for prenatal and even neonatal testing.

In addition, point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by yet other well-known techniques, e.g., direct DNA sequencing, single-strand
10 conformational polymorphism. See Orita et al., *Genomics*, 5, 874-879 (1989). For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed
15 by conventional procedures with radiolabeled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. The
20 presence of nucleotide repeats may correlate to a causative change in CIS activity or serve as marker for various polymorphisms.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in
25 electrophoretic mobility of DNA fragments in gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient
30 gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures. See, e.g., Myers et al., *Science*, 230, 1242 (1985). In addition, sequence alterations, in
35 particular small deletions, may be detected as changes in the migration pattern of DNA heteroduplexes in non-denaturing gel electrophoresis such as heteroduplex

electrophoresis. See, e.g., Nagamine et al., *Am. J. Hum. Genet.*, 45, 337-339 (1989). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or
5 the chemical cleavage method as disclosed by Cotton et al. in *Proc. Natl. Acad. Sci. USA*, 85, 4397-4401 (1985).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization (e.g., heteroduplex electroporation, see, White et al.,
10 *Genomics*, 12, 301-306 (1992), RNase protection (e.g., Myers et al., *Science*, 230, 1242 (1985)) chemical cleavage (e.g., Cotton et al., *Proc. Natl. Acad. Sci. USA*, 85, 4397-4401 (1985)), direct DNA sequencing, or
15 the use of restriction enzymes (e.g., restriction fragment length polymorphisms (RFLP) in which variations in the number and size of restriction fragments can indicate insertions, deletions, presence of nucleotide repeats and any other mutation which creates or destroys
20 an endonuclease restriction sequence). Southern blotting of genomic DNA may also be used to identify large (i.e., greater than 100 base pair) deletions and insertions.

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be
25 detected by *in situ* analysis. See, e.g., Keller et al., *DNA Probes*, 2nd Ed., Stockton Press, New York, N.Y., USA (1993). That is, DNA or RNA sequences in cells can be analyzed for mutations without isolation and/or immobilization onto a membrane. Fluorescence *in situ*
30 hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared. See, e.g., Trachuck et al., *Science*, 250, 559-562 (1990), and Trask et al., *Trends, Genet.*, 7, 149-154 (1991). Hence, by using nucleic acids based on
35 the structure of the CIS genes, one can develop diagnostic tests for genetic mutations.

In addition, some diseases are a result of, or are characterized by, changes in gene expression which can be detected by changes in the mRNA. Alternatively, the CIS gene can be used as a reference to identify

5 individuals expressing an increased or decreased level of CIS protein, e.g., by Northern blotting or *in situ* hybridization.

Defining appropriate hybridization conditions is within the skill of the art. See, e.g., "Current
10 Protocols in Mol. Biol." Vol. I & II, Wiley Interscience. Ausbel et al. (eds.) (1992). Probing technology is well known in the art and it is appreciated that the size of the probes can vary widely but it is preferred that the probe be at least 15
15 nucleotides in length. It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioisotopes, fluorescent dyes
20 or enzymes capable of catalyzing the formation of a detectable product. As a general rule, the more stringent the hybridization conditions the more closely related genes will be that are recovered.

The putative role of CIS in signal transduction of
25 the DNA synthesis pathway establishes yet another aspect of the invention which is gene therapy. "Gene therapy" means gene supplementation where an additional reference copy of a gene of interest is inserted into a patient's cells. As a result, the protein encoded by the
30 reference gene corrects the defect and permits the cells to function normally, thus alleviating disease symptoms. The reference copy would be a wild-type form of the CIS gene or a gene encoding a protein or peptide which modulates the activity of the endogenous CIS.

35 Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the

introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. A replication-deficient virus such as a modified retrovirus can be used to introduce the therapeutic CIS gene into such cells. For example, mouse Moloney leukemia virus (MLLV) is a well-known vector in clinical gene therapy trials. See, e.g., Boris-Lauerie et al., *Curr. Opin. Genet. Dev.*, 3, 102-109 (1993).

In contrast, *in vivo* gene therapy does not require isolation and purification of a patient's cells. The therapeutic gene is typically "packaged" for administration to a patient such as in liposomes or in a replication-deficient virus such as adenovirus as described by Berkner, K.L., in *Curr. Top. Microbiol. Immunol.*, 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in *Curr. Top. Microbiol. Immunol.*, 158, 97-129 (1992) and U.S. Patent No. 5,252,479. Another approach is administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue. Another approach is administration of "naked DNA" in which the therapeutic gene is introduced into the target tissue by microparticle bombardment using gold particles coated with the DNA.

Cell types useful for gene therapy of the present invention include lymphocytes, hepatocytes, myoblasts, fibroblasts, any cell of the eye such as retinal cells, epithelial and endothelial cells. Preferably the cells are T lymphocytes drawn from the patient to be treated, hepatocytes, any cell of the eye or respiratory or pulmonary epithelial cells. Transfection of pulmonary epithelial cells can occur via inhalation of a nebulized preparation of DNA vectors in liposomes, DNA-protein complexes or replication-deficient adenoviruses.

See, e.g., U.S. Patent No. 5,240,846.

Another aspect of the invention is transgenic, non-human mammals capable of expressing the polynucleotides of the invention in any cell. Transgenic, non-human animals may be obtained by transfecting appropriate
5 fertilized eggs or embryos of a host with the polynucleotides of the invention or with mutant forms found in human diseases. See, e.g., U.S. Patent Nos. 4,736,866; 5,175,385; 5,175,384 and 5,175,386. The resultant transgenic animal may be used as a model for
10 the study of CIS gene function. Particularly useful transgenic animals are those which display a detectable phenotype associated with the expression of the CIS protein. Drug development candidates may then be screened for their ability to reverse or exacerbate the
15 relevant phenotype.

The present invention will now be described with reference to the following specific, non-limiting examples.

20

Example 1

CIS full-length cDNA Cloning and Sequence Analysis

A search of a random cDNA sequence database consisting of short partial sequences known as expressed
25 sequence tags (ESTs) with SH2 domain encoding cDNA sequences using the BLASTX algorithm disclosed an EST which was homologous to the SH2 domain of the regulatory subunit of PI-3 kinase. This EST was originally isolated from a human endometrial tumor cDNA library.
30 Further searching revealed that the EST was homologous to the 3' end of a murine CIS cDNA sequence reported by A. Yoshimura et al., *supra*, (Genbank Accession No. D31943) (SEQ ID NO:3).

A circular 5'-rapid amplification of cDNA ends
35 (cRACE) protocol as described by Maruyama et al. in *Nucl. Acids Res.* 23, 3796-3797 (1995) was used to isolate the 5' cDNA end of the putative human CIS. One

hundred ng of human skeletal muscle polyA RNA (Clontech #6541-1) was reverse transcribed with MoMLV reverse transcriptase using a 5'-phosphorylated gene-specific primer GGCCACATAGTGCTGCACAA (SEQ ID NO:5). The single-stranded cDNA product was circularized by treatment with T4 RNA ligase. Two adjacent gene-specific primers GGAAGCTGGAGTCGGCATAAC (SEQ ID NO:6) and CTCCAAGCTGCTTGTCAGGC (SEQ ID NO:7), priming in opposite directions, were used to amplify by PCR a 0.5 kb fragment from the single-stranded circular cDNA template. PCR was conducted at 94°C for 20 s, 60°C to 40°C in 0.5°C increment/cycle for 30 s, 72°C for 2 min., for 40 cycles. The 0.5 kb 5' cRACE fragment was subcloned into pBluescript II and sequenced.

Sequence analysis revealed the fragment to be the 5'-end of the 3' clone of the putative human CIS, containing its remaining coding sequence, including the N-terminus. The encoded protein contains a central SH2 domain flanked by domains of unknown function.

A cDNA encoding an intact coding sequence was assembled. A 1.6 kb fragment was amplified from the 3' clone by PCR using the primer ACAGCACGCACCCCAGCTAC (SEQ ID NO:8) and the T7 primer. Similarly, a 0.5 kb fragment was amplified from the 5' cRACE product, isolated as described above, using the primers GGAAGCTGGAGTCGGCATAAC (SEQ ID NO:6) and CTCCAAGCTGCTTGTCAGGC (SEQ ID NO:7). These products were recombined by PCR in a second reaction containing each of the above PCR products and the primers GGAATTCCATGGTCCTCTGCGTTCAGGG (SEQ ID NO:9) and CCGTCGACGGTCAGAGCTGGAAGGGGTACT (SEQ ID NO:10). The PCR conditions for both sets of reactions were 94°C for 15 s, 55°C for 20 s, 72°C for 1 min., for 25 cycles. The 0.8 kb secondary PCR product was treated with EcoRI and SalI and subcloned into pBluescript II (Stratagene) and pGEX4T-3 (Pharmacia). High levels of CIS protein

expression in bacteria were observed with the pGEX4T-3 construct.

Independent confirmation of the existence of a mRNA corresponding to the full-length cDNA produced was carried out by RT-PCR. cDNA was prepared from 100 ng of human skeletal muscle polyA RNA (Clontech #6541-1) using random hexamer primers and MoMLV reverse transcriptase. One twentieth of the cDNA was used as template in a PCR reaction containing the primers ATGGTCCTCTGCGTTCAGGG (SEQ ID NO:11) and TCAGAGCTGGAAGGGGTACT (SEQ ID NO:12) and the expected 786 bp product was observed. The PCR conditions were 94°C for 15 s, 70°C to 50°C in 0.5°C increment/cycle for 20 s, 72°C for 2 min., for 40 cycles. Control reactions containing no template or containing the 0.8 kb recombined cDNA produced above gave either no PCR product or a 786 bp product, as expected.

Sequence analysis of the entire human CIS cDNA revealed a 774 nucleotide open reading frame (SEQ ID NO:1) encoding a 258 amino acid protein (SEQ ID NO:2) with a predicted molecular mass of 28.4 kDa, starting with an ATG at position 72 and terminating with a TGA at position 846 of SEQ ID NO:1. A proline-rich region is present at the C-terminus. The SH2 domain of human CIS is encoded by nucleotides 315 to 618 which correspond to residues 82 (Trp) to 183 (Val) in SEQ ID NO:2. GenBank searches using the BLASTX and BLASTP algorithms with the full-length DNA sequence or with the deduced amino acid sequence indicated that the human CIS SH2 domain was most homologous with the SH2 domain of the regulatory subunit of PI-3 kinase.

Alignment of the deduced amino acid sequence of human CIS (SEQ ID NO:2) with the murine CIS amino acid sequence (SEQ ID NO:4) was accomplished using the GAP algorithm. The overall amino acid identity was 90% with a one amino acid gap and is shown in Fig. 1 (top, human CIS; bottom, murine CIS).

Example 2Tissue Distribution of CIS

Northern blots of tissue mRNA were conducted to determine the tissue distribution of CIS gene

5 transcripts. The 1.8 kb insert of the 3' CIS clone was PCR amplified using T3 and T7 primers. Twenty five ng of the isolated PCR product was radiolabelled with [³²P]-dATP using a randomly primed labelling kit from Stratagene.

10 Membranes containing mRNA from multiple human tissues (Clontech #7760-1 and #7759-1) were hybridized with the probes and washed under high stringency conditions. Hybridized mRNA was visualized by exposing the membranes for six hours to a storage phosphor screen
15 (Molecular Dynamics). The results indicated that the 2.2 kb CIS transcript is largely ubiquitous and is expressed at variable levels in heart, placenta, lung, liver, muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, intestine, colon and peripheral
20 blood lymphocytes. Highest expression was observed in liver and ovary. The mRNA appears absent from brain.

The present invention may be embodied in other specific forms without departing from the spirit or
25 essential attributes thereof, and, accordingly, reference should be made to the appended claims, rather than to the foregoing specification, as indicating the scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: SmithKline Beecham Corporation and Harvard University
- (ii) TITLE OF THE INVENTION: HUMAN CIS PROTEIN
- (iii) NUMBER OF SEQUENCES: 12
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: SmithKline Beecham Corporation
 - (B) STREET: 709 Swedeland Road
 - (C) CITY: King of Prussia
 - (D) STATE: PA
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 19406-0939
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ Version 1.5
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US
 - (B) FILING DATE: 21-MAY-1996
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Baumeister, Kirk
 - (B) REGISTRATION NUMBER: 33,833
 - (C) REFERENCE/DOCKET NUMBER: P50486

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 610-270-5096
- (B) TELEFAX: 610-270-5090
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1374 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCCTGCACTG	CTGATACCCG	AAGCGACAGC	CCGATCCTGC	TCCCACTCCG	GAGTCGCCCG	60
TGCGCGGAGA	CATGGTCCCT	TGCGTTCAGG	GACCTCGTCC	TTTGCTGGCT	GTGGAGCGGA	120
CTGGGCAGCG	GCCCCGTGTT	GCCCCGTCCC	TGGAACGTGC	CAAGCCAGTC	ATGCAGCCCT	180
TGCCTGCTGG	GGCCTTCCTC	GAGGAGGTGG	CAGAGGGTAC	CCCAGCCCAG	ACAGAGAGTG	240
AGCCAAAGGT	GCTGGACCCA	GAGGAGGATC	TGCTGTGCAT	AGCCAAGACC	TTCTCCTACC	300
TTCGGGAATC	TGGCTGGTAT	TGGGGTTCCA	TTACGGCCAG	CGAGGCCCGA	CAACACCTGC	360
AGAAGATGCC	AGAAGGCACG	TTCTTAGTAC	GTGACAGCAC	GCACCCAGC	TACCTGTTCA	420
CGCTGTCACT	GAAAACCACT	CGTGGCCCCA	CCAATGTACG	CATTGAGTAT	GCCGACTCCA	480
GCTTCCGTCT	GGACTCCAAC	TGCTTGTTCA	GGCCACGCAT	CCTGGCCTTT	CCGGATGTGG	540
TCAGTCTTGT	GCAGCACTAT	GTGGCCTCCT	GCACTGCTGA	TACCCGAAGC	GACAGCCCCG	600
ATCCTGCTCC	CACCCCGGTC	CTGCCTATGC	CTAAGGAGGA	TGCGCCTAGT	GACCCAGCAC	660
TGCCTGCTCC	TCCACCAGCC	ACTGCTGTAC	ACCTAAAACT	GGTGCAGCCC	TTTGTACGCA	720
GAAGCAGTGC	CCGCAGCCTG	CAACACCTGT	GCCGCCTTGT	CATCAACCGT	CTGGTGGCCG	780
ACGTGGACTG	CCTGCCACTG	CCCCGGCGCA	TGGCCAACTA	CCTCCGACAG	TACCCCTTCC	840
AGCTCTGACT	GTACGGGGCA	ATCTGCCACC	CTCACCCAGT	CGCACCCCTG	AGGGGACATC	900
AGCCCCAGCT	GGACTTGGGC	CCCCACTGTC	CCTCCTCCAG	GCATCCTGGT	GCCTGCATAC	960
CTCTGGCAGC	TGGCCCAGGA	AGAGCCAGCA	AGAGCAAGGC	ATGGGAGAGG	GGAGGTGTCA	1020
CACAACTTGG	AGGTAAATGC	CCCCAGGCCG	CATGTGGCTT	CATTATACTG	AGCCATGTGT	1080
CAGAGGATGG	GGAGACAGGC	AGGACCTTGT	CTCACCTGTG	GGCTGGGCCC	AGACCTCCAC	1140
TCGATTGCCT	GCCCTGGTCA	CCTGAACGTG	ATGGGCACTC	TCAGCCCTGG	TTTTTCAATC	1200

```

CCCAGGGTCG GGTAGGACCC CTACTGCGCA GCCAGTCTCT TTTTCTGGGA GGATGACATG 1260
CAGCGGAAC T GAGATCGACA GTGTACTAGT GACCTCTTGT TGAGGGGTAA GCCAGGATAG 1320
GGGACTTGCA CAATCTATAC ACTATTTATT TATTTATTCT CCGTGGGGGT TGCA 1374

```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 258 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Val Leu Cys Val Gln Gly Pro Arg Pro Leu Leu Ala Val Glu Arg
 1              5              10              15
Thr Gly Gln Arg Pro Leu Trp Ala Pro Ser Leu Glu Leu Ser Lys Pro
      20              25              30
Val Met Gln Pro Leu Pro Ala Gly Ala Phe Leu Glu Glu Val Ala Glu
      35              40              45
Gly Thr Pro Ala Gln Thr Glu Ser Glu Pro Lys Val Leu Asp Pro Glu
      50              55              60
Glu Asp Leu Leu Cys Ile Ala Lys Thr Phe Ser Tyr Leu Arg Glu Ser
      65              70              75              80
Gly Trp Tyr Trp Gly Ser Ile Thr Ala Ser Glu Ala Arg Gln His Leu
      85              90              95
Gln Lys Met Pro Glu Gly Thr Phe Leu Val Arg Asp Ser Thr His Pro
      100             105             110
Ser Tyr Leu Phe Thr Leu Ser Val Lys Thr Thr Arg Gly Pro Thr Asn
      115             120             125
Val Arg Ile Glu Tyr Ala Asp Ser Ser Phe Arg Leu Asp Ser Asn Cys
      130             135             140
Leu Ser Arg Pro Arg Ile Leu Ala Phe Pro Asp Val Val Ser Leu Val
      145             150             155             160
Gln His Tyr Val Ala Ser Cys Thr Ala Asp Thr Arg Ser Asp Ser Pro
      165             170             175

```

Asp Pro Ala Pro Thr Pro Val Leu Pro Met Pro Lys Glu Asp Ala Pro
 180 185 190
 Ser Asp Pro Ala Leu Pro Ala Pro Pro Pro Ala Thr Ala Val His Leu
 195 200 205
 Lys Leu Val Gln Pro Phe Val Arg Arg Ser Ser Ala Arg Ser Leu Gln
 210 215 220
 His Leu Cys Arg Leu Val Ile Asn Arg Leu Val Ala Asp Val Asp Cys
 225 230 235 240
 Leu Pro Leu Pro Arg Arg Met Ala Asn Tyr Leu Arg Gln Tyr Pro Phe
 245 250 255
 Gln Leu

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2000 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTCCCCCTT	GTCCTTCCAA	GCTGTTGCA	CCACAGCCTT	TCAGTCCCTG	CTCGCCGCCC	60
GTGTGCCCCG	GGACCCTGAC	CTTCGCACCC	CTGGACCCAT	TGGCTCCTTT	CTCCTTCCAT	120
CCCGCCGAAC	TCCGACTCTC	GAGCCGCCGT	TGTCTCTGGG	ACATGGTCCT	CTGCGTACAG	180
GGATCTTGTC	CTTTGCTGGC	TGTGGAGCAA	ATTGGGCGGC	GGCCTCTGTG	GGCCCAGTCC	240
CTGGAGCTGC	CCGGGCCAGC	CATGCAGCCC	TTACCCACTG	GGGCATTCCC	AGAGGAAGTG	300
ACAGAGGAGA	CCCCTGTCCA	GGCAGAGAAT	GAACCGAAGG	TGCTAGACCC	TGAGGGGGAT	360
CTGCTGTGCA	TAGCCAAGAC	GTTCTCCTAC	CTTCGGGAAT	CTGGGTGGTA	CTGGGGTTCT	420
ATTACAGCCA	GCGAGGCCCG	GCAGCACCTA	CAGAAGATGC	CGGAGGGTAC	ATTCTTAGTT	480
CGAGACAGCA	CCCACCCAG	CTACCTGTTC	ACACTGTGTC	TCAAAACCAC	CCGTGGCCCC	540
ACCAACGTGC	GGATCGAGTA	CGCCGATTCT	AGCTTCCGGC	TGGACTCTAA	CTGCTTGTC	600
AGACCTCGAA	TCCTGGCCTT	CCCAGATGTG	GTCAGCCTTG	TGCAGCACTA	TGTGGCCTCC	660
TGTGCAGCTG	ACACCCGGAG	CGACAGCCCG	GATCCTGCTC	CCACCCAGC	CCTGCCTATG	720
TCTAAGCAAG	ATGCACCTAG	TGACTCGGTG	CTGCCTATCC	CCGTGGCTAC	TGCAGTGCAC	780

```

CTGAAACTGG TGCAGCCCTT TGTGCGCAGG AGCAGTGCCC GCAGCTTACA ACATCTGTGT      840
CGGCTAGTCA TCAACCGTCT GGTGGCCGAC GTGGACTGCT TACCCCTGCC CCGGCGTATG      900
GCCGACTACC TCCGACAGTA CCCCTTCCAA CTCTGACTGA GCCAGGCACC CTGCTCTGCC      960
TCACACAGTC ACATCCTGGA GGGAACACAG TCCCCAGCTG GACTTGGGGT TCTGCTGTCC     1020
TTTCTTCAGT CATCCTGGTG CCTGCATGCA TGTGACAGCT GGACCAGAGA ATGCCAGCAA     1080
GAACAAGGCA GGTGGAGGAG GGATTGTCAC ACAACTCTGA GGTCAACGCC TCTAGGTACA     1140
ATATGGCTCT TTGTGGTGAG CCATGTATCA GAGCGAGACA GGCAGGACCT CGTCTCTCCA     1200
CAGAGGCTGG ACCTAGGTCT CCACTCACTT GCCTGCCCTT GCCACCTGAA CTGTGTCTAT     1260
TCTCCCAGCC CTGGTTTCTC AGTCTGCTGA GTAGGGCAGG CCCCTACCC ATGTATAGAA     1320
TAGCGAGCCT GTTTCTGGGA GAATATCAGC CAGAGGTTGA TCATGCCAAG GCCCCTTATG     1380
GGGACGCAGA CTGGGCTAGG GGACTIONACA GTTATACAGT ATTTATTTAT TTATTCTCCT     1440
TGCAGGGGTT GGGGGTGGAA TGATGGCGTG AGCCATCCCA CTTCTCTGCC CTGTGCTCTG     1500
GGTGGTCCAG AGACCCCCAG GTCTGGTTCT TCCCTGTGGA GACCCCCATC CAAAACATT     1560
GTTGGGCCCCA AAGTAGTCTC GAATGTCTCT GGCCCATCCA CCTGCGTATG GATGTGCCCA     1620
CTTTTTTCTC CCAAGCCTCT TTTGGGAGGC TGGGTGGCCA GACAGACAGG AGCCAGAAAC     1680
ACAAGGGCTC CCACTCTTCT CCTCACAGGG CAGCACCATG GCTTCATAGA GCTGGCTTCT     1740
CTATGTTGTG CCCCACCTCA CCCCCCTGCC GAGGGGCGTG TGCTGGGTCG GGAAGTGGAT     1800
GCTTATCCAA GGGCCGAGA GTAGCTCCC TTGTGTCCGT TTCCTGCCTA GGAAGTTGCC     1860
TGCACGTGAG AGAGGGAGAA ATACATACAC ACCTAACAAG ACTTTAGAAA ACAAGTGTTA     1920
GAACACAAGA ACCAGTTTGG GAGTTTTTCT TCCACTGATT TTTTCTGTA ATGATAATAA     1980
AATTATGCCT TCCACTTATG                                     2000

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Val Leu Cys Val Gln Gly Ser Cys Pro Leu Leu Ala Val Glu Gln
 1              5              10              15
Ile Gly Arg Arg Pro Leu Trp Ala Gln Ser Leu Glu Leu Pro Gly Pro
      20              25              30

```

Ala Met Gln Pro Leu Pro Thr Gly Ala Phe Pro Glu Glu Val Thr Glu
 35 40 45
 Glu Thr Pro Val Gln Ala Glu Asn Glu Pro Lys Val Leu Asp Pro Glu
 50 55 60
 Gly Asp Leu Leu Cys Ile Ala Lys Thr Phe Ser Tyr Leu Arg Glu Ser
 65 70 75 80
 Gly Trp Tyr Trp Gly Ser Ile Thr Ala Ser Glu Ala Arg Gln His Leu
 85 90 95
 Gln Lys Met Pro Glu Gly Thr Phe Leu Val Arg Asp Ser Thr His Pro
 100 105 110
 Ser Tyr Leu Phe Thr Leu Ser Val Lys Thr Thr Arg Gly Pro Thr Asn
 115 120 125
 Val Arg Ile Glu Tyr Ala Asp Ser Ser Phe Arg Leu Asp Ser Asn Cys
 130 135 140
 Leu Ser Arg Pro Arg Ile Leu Ala Phe Pro Asp Val Val Ser Leu Val
 145 150 155 160
 Gln His Tyr Val Ala Ser Cys Ala Ala Asp Thr Arg Ser Asp Ser Pro
 165 170 175
 Asp Pro Ala Pro Thr Pro Ala Leu Pro Met Ser Lys Gln Asp Ala Pro
 180 185 190
 Ser Asp Ser Val Leu Pro Ile Pro Val Ala Thr Ala Val His Leu Lys
 195 200 205
 Leu Val Gln Pro Phe Val Arg Arg Ser Ser Ala Arg Ser Leu Gln His
 210 215 220
 Leu Cys Arg Leu Val Ile Asn Arg Leu Val Ala Asp Val Asp Cys Leu
 225 230 235 240
 Pro Leu Pro Arg Arg Met Ala Asp Tyr Leu Arg Gln Tyr Pro Phe Gln
 245 250 255
 Leu

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGCCACATAG TGCTGCACAA

20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGAAGCTGGA GTCGGCATA

20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTCCAAGTGC TTGTCCAGGC

20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACAGCACGCA CCCCAGCTAC

20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGAATTCCAT GGTCTCTGCT GTTCAGGG

28

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid

40

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCGTCGACGG TCAGAGCTGG AAGGGGTACT

30

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTISENSE: NO
 - (v) FRAGMENT TYPE:
 - (vi) ORIGINAL SOURCE:
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGGTCCTCT GCGTTCAGGG

20

- (2) INFORMATION FOR SEQ ID NO:12:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCAGAGCTGG AAGGGGTACT

20

CLAIMS

1. An isolated polynucleotide selected from the group consisting of:
 - (a) a polynucleotide encoding human CIS having the nucleotide sequence as set forth in SEQ ID NO: 1 from nucleotide 72 to 846;
 - (b) a polynucleotide capable of hybridizing to the complement of a polynucleotide according to (a) under moderately stringent hybridization conditions and which encodes a functional human CIS; and
 - (c) a degenerate polynucleotide according to (a) or (b).
2. An isolated polynucleotide having the nucleotide sequence as set forth in SEQ ID NO:1.
3. A functional polypeptide encoded by the polynucleotide of claim 1.
4. The functional polypeptide of claim 3 which is human CIS having the amino acid sequence set forth in SEQ ID NO:2.
5. The polynucleotide of claim 1 which is DNA.
6. The polynucleotide of claim 5 which is genomic DNA.
7. The polynucleotide of claim 1 which is RNA.
8. A vector comprising the DNA of claim 5.
9. A recombinant host cell comprising the vector of claim 8.
10. A method for preparing essentially pure human CIS protein comprising culturing the recombinant host cell of claim 9 under conditions promoting expression of the protein and recovering the expressed protein.
11. Human CIS produced by the process of claim 10.
12. An antisense oligonucleotide comprising a sequence which is capable of binding to the polynucleotide of claim 1.
13. A modulator of the polypeptide of claim 3.
14. The modulator of claim 13 which is a peptide.

15. The modulator of claim 13 which is a small organic molecule.

16. The small organic molecule of claim 15 which is a peptidomimetic.

17. A method for assaying a medium for the presence of a substance that modulates CIS activity by affecting the binding of CIS to cellular binding partners comprising the steps of:

(a) providing a CIS protein having the amino acid sequence of CIS (SEQ ID NO:2) or a functional derivative thereof and a cellular binding partner or synthetic analog thereof;

(b) incubating with a test substance which is suspected of modulating CIS activity under conditions which permit the formation of a CIS protein/cellular binding partner complex;

(c) assaying for the presence of the complex, free CIS protein or free cellular binding partner; and

(d) comparing to a control to determine the effect of the substance.

18. CIS protein modulating compounds identified by the method of claim 17.

19. A method for the treatment of a patient having need to modulate CIS activity comprising administering to the patient a therapeutically effective amount of the modulating compound of claims 18.

20. A pharmaceutical composition comprising the modulating compound of claim 18 and a pharmaceutically acceptable carrier.

21. A method for assaying for the presence of a substance that modulates CIS activity by direct binding to CIS protein comprising the steps of:

(a) providing a labelled CIS protein having the amino acid sequence of CIS (SEQ ID NO:2) or a functional derivative thereof

(b) providing solid support-associated modulator candidates;

(c) incubating a mixture of the labelled CIS protein with the support-associated modulator candidates under conditions which can permit the formation of a CIS protein/modulator candidate complex;

(d) separating the solid support from free soluble labelled CIS protein;

(e) assaying for the presence of solid support-associated labelled protein;

(f) isolating the solid support complexed with labelled CIS protein; and

(g) identifying the modulator candidate.

22. CIS protein modulating compounds identified by the method of claim 21.

23. A method for the treatment of a patient having need to modulate CIS activity comprising administering to the patient a therapeutically effective amount of the modulating compound of claim 21.

24. A pharmaceutical composition comprising the modulating compound of claim 21 and a pharmaceutically acceptable carrier.

25. A method of diagnosing conditions associated with CIS protein deficiency which comprises:

(a) isolating a polynucleotide sample from an individual;

(b) assaying the polynucleotide sample and a polynucleotide encoding CIS having the nucleotide sequence as set forth in SEQ ID NO:1 from nucleotide 72 to 846; and

(c) comparing differences between the polynucleotide sample and the CIS polynucleotide, wherein any differences indicate mutations in the CIS gene.

26. A method of treating conditions which are related to insufficient CIS protein function which comprises:

(a) isolating cells from a patient deficient in CIS protein function;

(b) altering the cells by transfecting the polynucleotide of claim 1 into the cells wherein a CIS protein is expressed; and

(c) introducing the cells back to the patient to alleviate the condition.

27. A method of treating conditions which are related to insufficient CIS protein function which comprises administering the polynucleotide of claim 1 to a patient deficient in CIS protein function wherein a CIS protein is expressed and alleviates the condition.

28. A transgenic non-human animal capable of expressing in any cell thereof the DNA of claim 5.


```

1  MVLCVQGPRLPLLAVERTGQRPLWAPSLELSKPVMQPLPAGAFLEEVAEGT  50
   |||||.|. |||||. .|. |||||. |||. |. |||||. ||| |||. :|
5  1  MVLCVQGSCPLLAVEQIGRRPLWAQSLPLPGPAMQPLPTGAFPEEVTEET  50
   .
   .
51 PAQTESEPKVLDPEEDLLCIAKTF SYLRESGWYWG SITASEARQHLQKMP  100
   |.|.|. ||||| |: ||||| ||||| ||||| ||||| ||||| |||||
10 51 PVQAENEPKVLDPEGDLLCIAKTF SYLRESGWYWG SITASEARQHLQKMP  100
   .
   .
101 EGTFLVRDSTHPSYLF TSLVKTTRGPTNVRIEYADSSFR LDSNCLSRPRI  150
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 EGTFLVRDSTHPSYLF TSLVKTTRGPTNVRIEYADSSFR LDSNCLSRPRI  150
   .
   .
15 151 LAFPDVVSLVQHYVASCTADTRSDSPDPAPT PVLMPKEDAPSDPALPAP  200
   ||||| ||||| ||||| . ||||| ||||| ||||| . |||. :| |||||. .||
151 LAFPDVVSLVQHYVASCAADTRSDSPDPAPT PALPMSKQDAPSDSVLP I.  199
   .
   .
20 201 PPATAVHLKLVQPFVRRSSARS LQHL CRLVINRLVADV DCLPLPRRMANY  250
   |. ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :|
200 PVATAVHLKLVQPFVRRSSARS LQHL CRLVINRLVADV DCLPLPRRMADY  249
   .
   .
25 251 LRQYPFQL 258
   |||||
250 LRQYPFQL 257

```

FIGURE 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/07477

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : 536/23.1, 23.5, 24.31; 530/350; 435/69.1, 252.3, 320.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.5, 24.31; 530/350; 435/69.1, 252.3, 320.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN, BIOSIS, SCISEARCH, MEDLINE, DIALOG, DERWENT
search terms: CIS, SH2 protein, human

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YOSHIMURA et al. A novel cytokine-inducible gene CIS encodes an SH2-containing protein that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin receptors. The EMBO Journal. 1995, Vol. 14, No. 12, pages 2816-2826, see entire document, especially page 2818, Figure 2.	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be part of particular relevance

E earlier document published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

* * document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

G

document member of the same patent family

Date of the actual completion of the international search

27 AUGUST 1996

Date of mailing of the international search report

04 OCT 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DARYL A. BASHAM

Telephone No.

(703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/07477

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-12

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/07477

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C07H 21/02, 21/04; C07K 1/00, 14/435; C12N 15/12, 15/63, 15/64; C12P 21/00

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-12, drawn to an isolated polynucleotide, an isolated polypeptide, DNA, RNA, a vector, a host cell containing a vector, a method for preparing an essentially pure protein and an antisense oligonucleotide.
- II. Claims 13-16, 18, 20, 22 and 24, drawn to modulators of CIS protein activity.
- III. Claims 17 and 21, drawn to methods for assaying for modulators of CIS protein activity.
- IV. Claims 19 and 23, drawn to methods of treatment of a patient having need to modulate CIS activity.
- V. Claim 25, drawn to method of diagnosing condition of deficiency of CIS comprising determining a mutation in a nucleotide sequence.
- VI. Claims 26 and 27, drawn to gene therapy.
- VII. Claim 28, drawn to a transgenic animal.

The polynucleotide and polypeptide compositions of Group I have materially different chemical structures and biological functions from the modulators of Group II. The special features by which the polynucleotide and polypeptide compositions are defined distinguish them from the special technical features which define the modulators of Group II. The host cell of Group I is materially different from the transgenic animal of Group VII. The special technical features which distinguish the host cell of Group I from the transgenic animal of Group VII are the process steps required. Because the process steps required to create the transgenic animal of Group VII are not the same as the process steps required to make the host cells, unity of invention is lacking. The method of each group is materially different from the method of any other group because each is practiced with materially different process steps; the process steps are the special technical feature which distinguish each method from the others. Because the process steps do not share the same or corresponding special technical feature, unity of invention is lacking. The claims are not so linked by a special technical feature within the meaning of the PCT Rule 13.2 so as to form a single inventive concept.